Risk Stratification in Severe Sepsis: Organ Failure Scores, PIRO or Both?

R.P. Moreno, P. Metnitz, and P. Bauer

“It's more important to know what sort of person this disease has, than what sort of disease this person has”

William Osler 1849 – 1919

Introduction

The use of all-cause hospital mortality as the sole or major endpoint for the evaluation of clinical trials in intensive care was challenged in the mid 1980s, in the aftermath of a very long series of negative clinical trials in patients with sepsis, severe sepsis, and septic shock [1]. This outcome measure, until then viewed as the golden standard in clinical trials in intensive care, is, beyond any doubt, a very relevant endpoint both for researchers and for clinicians. Its use has been contested because hospital policy can and does change the location of deaths (e.g., discharging patients to the ward to die) and mortality rates can, therefore, be significantly underestimated in hospitals that discharge patients very early in the course of their disease to other facilities [2].

Moreover, the absence of stratification in the process of selecting two groups of patients, one assigned to receive the intervention under study and the other assigned to receive the placebo, has been criticized. The absence of stratification according to patient demographic or biological characteristics before randomization can often result in unbalanced groups and in the presence of confounding; an impossibly high number of patients would therefore need to be enrolled in order to demonstrate a significant difference between patient groups. Also, interactions between certain patient characteristics at baseline and the effect of treatment can be obscured, as happened in the Monoclonal Anti-TNF: A Randomized Clinical Sepsis (MONARCS) trial, in which the administration of afelimomab was able to lower circulating levels of tumor necrosis factor (TNF) and interleukin (IL)-6, accelerate the resolution of organ dysfunction, and reduce 28-day all-cause mortality but only in patients with elevated IL-6 levels at baseline [3].

For all these reasons, some investigators proposed at that time that patients should be stratified according to several factors, not all of them included in the American College of Chest Physicians (ACCP)/Society of Critical Care Medicine (SCCM) definitions of sepsis or sepsis syndrome. These included the use of a validated scoring system for organ dysfunction that could be incorporated into sepsis studies, such that major morbidities (and not only 28-day all-cause mortality) should be considered as primary end points [4].

Soon after the publication of the Protein C Worldwide Evaluation in Severe Sepsis (PROWESS) trial by Gordon Bernard and co-workers [5], the presence of several confounders and effect modifiers became evident. The most important of these was the baseline severity of illness and the site of infection [5, 6]. These potential confounders, in addition to later publications with discrepant results either in controlled [7] or uncontrolled settings [8, 9], have raised such a debate [10], that the
drug is now being assessed in a risk-stratified population, the so-called PROWESS-SHOCK study [11].

Results of the Corticosteroid Therapy of Septic Shock (CORTICUS) study, comparing the use of hydrocortisone with placebo in patients with septic shock, were published in 2008 by Charles Sprung and co-workers [12]. Again, baseline severity of illness and possibly other baseline- and infection-related factors played an important role in the interpretation of results. These factors could be responsible for the negative results of the intervention on all-cause 28-day mortality, compared to another study with almost the same design, but done in a cohort of more severely ill patients [13]. This effect is even more striking, because hydrocortisone, despite having no effect on all-cause 28-day mortality, did reduce the length of shock and the severity of multiple organ failure (MOF), specifically the amount of cardiovascular failure [14].

Authors, such as Petros and colleagues, questioned almost 15 years ago the adequacy of all-cause mortality as an endpoint [15]. A meaningful endpoint can only be chosen when a direct relation between an event and its consequences is known. In the case of sepsis (and MOF), our knowledge is very limited and no direct relationship can be established. Moreover, the use of this outcome variable without prior patient stratification implies the need for large samples to have a fair chance to achieve an adequate balance between groups, creating problems in reliability of data collection, heterogeneity of enrolled patients, duration of the trial, and associated costs. Patients in intensive care units (ICUs), even with very strict inclusion criteria for sepsis or septic shock, do not constitute a homogeneous sample. Patients have different ages, chronic illnesses (chronic health, comorbidities), diagnoses, time-courses, sites of infection and invading microorganisms, and are enrolled in trials with different degrees of physiologic dysfunction resulting in a large dispersion of mortality risks, as demonstrated in the mid 1990s [16, 17]. Several methods have been proposed by different authors to deal with such huge variation [17–19], but these usually lead to complex, extensive (and expensive) data collection systems and are thus not feasible.

Two approaches have been proposed and tested to cope in a more intuitive way with this complex problem of patient selection and stratification. This includes the development and validation of organ dysfunction scores (based on the presence and severity of organ dysfunction/failure in some organs and systems) and the so-called PIRO-scores.

Organ Dysfunction/Failure Scores

The awareness of the importance of evaluating and considering MOF as an important confounder and/or effect modifier in the evaluation of patients with sepsis lead the Working Group on Sepsis Related Problems of the European Society of Intensive Care Medicine (ESICM), under the leadership of Professor Jean-Louis Vincent, to organize in December 1994, in Paris, a consensus meeting to create the so-called sepsis-related (later renamed sequential) organ failure assessment (SOFA) score [20]. The rationale behind this decision was the need to find an objective and simple way to describe individual organ dysfunction/failure in a continuous form, from mild dysfunction to severe failure, that could be used over time to measure the evolution of individual (or aggregated) organ dysfunction in clinical trials on sepsis or for the clinician at the bedside. A retrospective evaluation of the application of this score to the first 24 hours in the ICU on 1,643 patients with early sepsis on an international